

A Greek Corpus of Aphasic Discourse: Collection, Transcription, and Annotation Specifications

Spyridoula Varlokosta¹, Spyridoula Stamouli^{1,2}, Athanasios Karasimos^{1,3}, Georgios Markopoulos¹, Maria Kakavoulia⁴, Michaela Nerantzini^{1,5}, Aikaterini Pantoula¹, Valantis Fyndanis^{1,6}, Alexandra Economou¹, Athanassios Protopapas¹

¹National and Kapodistrian University of Athens

²Institute for Language and Speech Processing / “Athena” Research Center

³Academy of Athens

⁴Panteion University of Social and Political Sciences

⁵Northwestern University

⁶University of Oslo

Address: University of Athens, Philology/Linguistics, Panepistimioupoli Zografou, Athens, 15784 Greece.

E-mail: svarlokosta@phil.uoa.gr, pstam@ilsp.gr, akarasimos@academyofathens.gr, gmarkop@phil.uoa.gr, markak@panteion.gr, nmixaela@gmail.com, aikaterini.pantoula@gmail.com, valantis.fyndanis@iln.uio.no, aoikono@psych.uoa.gr, aprotopapas@phs.uoa.gr

Abstract

In this paper, the process of designing an annotated Greek Corpus of Aphasic Discourse (GREECAD) is presented. Given that resources of this kind are quite limited, a major aim of the GREECAD was to provide a set of specifications which could serve as a methodological basis for the development of other relevant corpora, and, therefore, to contribute to the future research in this area. The GREECAD was developed with the following requirements: a) to include a rather homogeneous sample of Greek as spoken by individuals with aphasia; b) to document speech samples with rich metadata, which include demographic information, as well as detailed information on the patients' medical record and neuropsychological evaluation; c) to provide annotated speech samples, which encode information at the micro-linguistic (words, POS, grammatical errors, clause types, etc.) and discourse level (narrative structure elements, main events, evaluation devices, etc.). In terms of the design of the GREECAD, the basic requirements regarding data collection, metadata, transcription, and annotation procedures were set. The discourse samples were transcribed and annotated with the ELAN tool. To ensure accurate and consistent annotation, a Transcription and Annotation Guide was compiled, which includes detailed guidelines regarding all aspects of the transcription and annotation procedure.

Keywords: aphasia, aphasic discourse, annotated corpus

1. Introduction

Aphasia is defined as a language disorder following a focal damage to the left cerebral hemisphere caused either by a cerebral vascular accident (CVA), a traumatic brain injury (TBI), an infection, such as encephalitis, or as the result of the existence or the removal of a brain tumor (De Roo, 1999: 1; Mesulam, 2000: 296). Aphasia is typically restricted to language impairments in the absence of any other general cognitive impairment or dementia (Oblor & Gjerlow, 1999: 38). Deficits in aphasia can potentially affect speech production and comprehension in both oral and written language forms, and at all linguistic levels (i.e., phonological, morphological, syntactic, and semantic), to varying degrees depending on the site and the severity of the brain injury (Harley, 2001: 23); from mild, to moderate and severe disorders.

Although in the aphasiological literature many different types of aphasia have been described, the most widespread classification identifies two basic categories: non-fluent aphasia or Broca's aphasia and fluent aphasia or Wernicke's aphasia, each one of which has been associated with different neurological characteristics, as for the locus and the extent of the lesion, and different linguistic characteristics.

Studies on speakers with aphasia conducted over the past 40 years have emphasized the clinical importance of the study of discourse production (e.g. Berko-Gleason et al., 1980; Nicholas & Brookshire, 1993; Olness & Ulatowska,

2011; Saffran, Berndt & Schwartz, 1989; Ulatowska, North & Macaluso-Haynes, 1981; Ulatowska et al., 1983; Vermeulen, Bastiaanse & van Wagensingen, 1989; see Armstrong, 2000, for an overview of the literature). Since people with aphasia experience particular difficulties in their everyday communication, the study of their abilities at the discourse level is considered as a natural and objective method for assessing the communicative effectiveness of these individuals in their everyday life. More specifically, the study of discourse production can contribute to the diagnosis of the type of aphasia, to a more accurate identification of the communication impairments of patients, to the design of a more effective treatment as well as to the evaluation of patients' response to treatment (Wright, 2011).

Despite the fact that there is a large body of literature on the characteristics of aphasic discourse in many languages, which includes studies following different methodological approaches, theoretical frameworks, and analytical perspectives, there is a considerable lack of available resources to allow the systematic study of aphasic discourse in a comparable and replicable way across languages. The available corpora of aphasic discourse -constructed with the use of corpus linguistic techniques and providing systematic methods for the transcription, annotation, and analysis- are the Corpus of Dutch Aphasic Speech (CoDAS Westerhout & Monachesi, 2006), the Cambridge Cookie-Theft Corpus (Williams et al., 2010),

and the AphasiaBank (MacWhinney et al., 2011, 2012). Each one of them has contributed from a different perspective to the process of enriching the existing methods and data for the study of aphasic discourse, and, consequently, to the advancement of research in this area. CoDAS comprises a pilot study of six aphasic speakers with two levels of annotation, an orthographic-phonetic transcription and a Part-Of-Speech (POS) tagging. The Cambridge Cookie-Theft Corpus contains transcriptions of spontaneous speech and single-picture descriptions elicited with the cookie-theft picture. The study includes data from approximately 87 brain-damaged patients in comparison to a group of 227 healthy individuals. A total of 1331 utterances are time-stamped and annotated on the phonological level following an XML-based TEI schema. AphasiaBank is a multimedia database with video and speech annotated transcriptions of approximately 180 speakers with aphasia and 140 non brain-damaged controls in a variety of communicative tasks and interactions. The transcriptions are based on the CHAT format and coded for analysis with specific CLAN programs. A multi-level annotation produces a language profile that includes word-level and utterance-level morphosyntactic errors.

The Greek Corpus of Aphasic Discourse (GREECAD) is the outcome of a research action under the large scale multidisciplinary project “THALES-Levels of impairment in Greek aphasia: relationship with processing deficits, brain region, and therapeutic implications”. The aim of this research action was the collection, annotation, documentation, and linguistic analysis of spoken discourse of Greek speakers with aphasia.

The development of the GREECAD had to meet the following requirements: a) to include a rather homogeneous sample of Greek as spoken by individuals with mild non-fluent aphasia; b) to document speech samples with rich metadata, which include demographic information, information on the patients’ medical record, as well as their speech and language therapy and neuropsychological evaluation; c) to provide annotated speech samples which encode properties of speech as well as linguistic information at the micro-linguistic (words, POS, grammatical, semantic, and phonological errors, clause types, etc.) and discourse level (narrative structure units, main events, evaluation devices, etc.).

In this paper, the process of designing the GREECAD is presented, regarding data collection, transcription, and annotation of speech samples. Given that resources of this kind are quite limited, a major aim of the development of the GREECAD was to provide a set of specifications which could serve as a methodological basis for the development of other relevant corpora, and, therefore, to contribute to future research in this area.

2. Data Collection

Among the various discourse types that have been studied, narrative discourse has attracted more attention in aphasia research, mainly because the abstract narrative schema provides an objective framework for the analysis of

speakers’ productions and their comparison to healthy controls. Therefore, a protocol of four narrative tasks (Kakavoulia et al., 2014) was developed to elicit spoken discourse samples from Greek speakers with aphasia. Previous research (Doyle et al., 1998) shows that the discourse produced by speakers with aphasia is influenced by the characteristics of elicitation tasks, such as the type of stimuli and the modality of presentation, as well as by the cognitive and linguistic requirements of the tasks, depending on the particular clinical characteristics of each individual. Thus, it was decided that the protocol should include different narrative genres (personal narrative, third person narrative, fairy tale, etc.) and different elicitation techniques (McNeil et al., 2007; Menn, Ramsberger & Helm-Estabrooks, 1994, Nicholas & Brookshire, 1995; Ulatowska et al., 1983), which provide different degrees and types of support to the participants in order to compensate for the cognitive and linguistic demands of each task. Personal narratives were chosen because they elicit more natural speech data characterized by extensive use of evaluative devices (Ulatowska et al., 2006). More constrained elicitation tasks, such as story retelling and picture elicitation, were also employed to ensure more controlled discourse samples. More specifically, the protocol includes the following tasks:

Task 1: Unaided production of a personal narrative (“stroke story”). The individuals with aphasia narrate the incident of their stroke story, while the control group (people who have suffered a heart attack, see Section 3, Participants) narrate the heart attack incident.

Task 2: Production of an unknown story based on a 6-picture series (“the party”). The participant narrates a short, simple story shown in the pictures presented to her/him by the researcher. Linguistic demands are high, since the participant has to generate the story events and the narrative structure from the pictures, but there are no memory requirements.

Task 3: Retelling of an unknown story, aided by a 5-picture series (“the ring”). The participant listens to a recorded story, which has the structure of a traditional fairy tale. The story is quite lengthy; it has many episodes and a complex plot, characteristics which increase the linguistic and cognitive demands of the task. At the same time, five pictures depicting important events of the story are presented to her/him. After listening to the story, the participant has to retell it using the pictures. Visual support is expected to compensate for the increased linguistic and cognitive demands of the task.

Task 4: Familiar story retelling (“hare and tortoise” Aesop’s fable). The participant listens to a recorded narration of the fable and afterwards she/he has to retell the story to the researcher. No visual support is used. Memory load is increased, since the participant has to retain story elements and their temporal order. However, this demand is compensated by the fact that the story is already familiar to the speaker.

3. Participants

The GREECAD contains spoken discourse samples

elicited from Greek-speaking individuals with mild non-fluent aphasia and controls matched for age and level of education to the aphasic speakers (Table 1).

	N	Age range (years)	Sex		Education	
			M	F	Years	N
Speakers with aphasia	18	39-67	15	3	6	3
					9	1
					12	5
					over 12	9
Control group	7	43-71	7	0	9	2
					12	2
					over 12	3

Table 1: Characteristics of participants

The control group comprises individuals who have suffered a heart attack. The choice of these individuals as a control group was made to ensure the comparability of the personal narrative samples (Task 1) in terms of their textual characteristics. Therefore, in accordance to the stroke story production by speakers with aphasia, the control group participants narrated their “heart attack story”. Heart attack is a similar traumatic experience to the stroke, with a comparable informational content and event sequence (initial symptoms, reaction from the part of the patient and relatives, medical diagnosis and intervention, outcomes). Although a few differences in the vocabulary were expected between the two versions of the personal narrative, mainly with respect to specific symptoms or medical treatment, their overall linguistic, structural and

informational similarities were considered as more useful for comparison between the two groups.

Ethical approval was obtained by the Ethics Committee of the hospitals, medical and rehabilitation centres involved in the project. Patients received written information about the study and were asked to provide full informed consent.

4. The GREECAD Corpus

The spoken discourse samples of speakers with aphasia and those of the control group were manually transcribed and annotated. The result of this process was the compilation of the GREECAD. Speakers with aphasia are currently represented in the corpus with 72 transcripts, while the control group with 28 transcripts. Table 2 shows the total number (N) of transcripts, tokens and clauses per group. Specific measurements on tokens and clauses, besides total count, include the statistical mean, as well as the minimum and maximum value per speaker in the corresponding group. The corpus is still being enriched with new data collected from individuals with aphasia and controls.

The discourse samples are documented with rich metadata which include demographic information about the participants, as well as detailed information on the patients’ medical record, including the type of aphasia, and their speech and language therapy and neuropsychological evaluation (e.g. their scores on the Boston Diagnostic Aphasia Examination, Greek version: Paphanassiou et al., 2008, and the Boston Naming Test, Greek version: Simos, Kasselimis & Mouzaki, 2011).

	Transcripts	Tokens				Clauses			
		N	N	Mean	Min	Max	N	Mean	Min
Speakers with aphasia	72	4643	64.5	14	117	1158	16.1	4	36
Control group	28	4006	143.1	68	208	871	31.1	12	56
Total	100	8649				2039			

Table 2: The GREECAD Corpus

5. Transcription and Annotation

Discourse samples were manually transcribed and annotated using the ELAN transcription and annotation tool (Wittenburg et al., 2006). Transcription was orthographic using the Greek alphabet. A transcription protocol was designed to encode the necessary information for linguistic analysis, excluding detailed phonological information. Specific conventions were used for unintelligible words and neologisms, while no special symbols were used. The transcripts were time-aligned with the audio files at utterance, clause, and word level. Annotation was carried out to encode linguistic information of the patients’ discourse at various levels. A

structured, multi-tiered annotation scheme was designed in order to include all the parameters of spoken discourse under investigation. These parameters include speech and non-speech events (e.g. vowel and consonant lengthening, pauses, filled gaps, laughter, etc.), micro-linguistic features (words, POS, grammatical, semantic, and phonological errors, clause types, etc.), as well as discourse features (narrative structure units, main events, evaluation devices). The annotated corpus is available in XML / EAF format, which allows the future analysis of data with automatic computational linguistic techniques. It was based on the Formal Framework for Linguistic Annotations (Bird & Liberman, 1999; Ide & Suderman, 2007, 2014) and the template is governed by token-based, type-based, and graph-based hierarchy.

For ensuring accurate and consistent transcription and annotation, a set of explicit and clear procedures was established, together with detailed guidelines for the annotators, which comprised a Transcription and Annotation Guide (Varlokosta et al., 2013). The annotators were graduate or postgraduate students of Linguistics. They were divided into small groups of 2-3 annotators. Each group annotated only certain tiers in each narrative, according to their specialisation, experience and interests, and not all tiers. Annotation tiers were assigned to groups as follows:

Group 1: Transcription (Researcher – Patient)

Group 2: Processed Transcription – Utterances (limits) – Clauses (limits)

Group 3: Events

Group 4: Clauses (tagging: types, grammaticality, completeness)

Group 5: Words (limits, POS tagging, counting)

Group 6: Errors (tagging: phonological, morphosyntactic, lexical/semantic errors, paraphrases)

Group 7: Reformulations

Group 8: Narrative annotation (narrative structure elements, main events, evaluation devices)

It should be noted that in some cases there was a single annotator in each group (e.g. group 1, 2, 3, 7). A two-person leading team was appointed to train and coordinate the groups of annotators. The annotation leaders were experts in Corpus Linguistics, experienced in data collection and processing with the use of the ELAN tool. This team trained each individual group in the annotation of the specific tiers they were assigned to. After training, a pilot phase was carried out, including two phases: a) initially, annotators were given a file in which they annotated their tiers in collaboration with one of the trainers, who helped them and resolved any query on the spot; b) subsequently, annotators annotated another file on their own, using the Transcription and Annotation Guide. Their annotations were checked by their trainer, who gave feedback regarding problematic issues. Phase b was repeated as many times as needed to ensure agreement of an acceptable level between the annotator and the trainer (above 90%). It should be noted that some annotations were more difficult than others (e.g. setting the utterance limits or tagging error types at word level), which led to more repetitions of phase b until the annotator and the trainer reached an agreement. This procedure highlighted the need for more explicit and detailed criteria for annotating these particularly difficult tiers. The pilot phase was carried out with each new annotator who entered a group. Before marking each file as “complete”, a checking phase was carried out, during which all annotations were checked by the leading team, who made the necessary corrections. Each member of the leading team was responsible for checking specific tiers. During the checking phase, the leading team provided feedback to the annotators, including new guidelines, if needed. All the new instructions and modifications regarding the annotation scheme, the guidelines, as well as specific annotation criteria that came up during the pilot and the checking phase were integrated into the Transcription and Annotation Guide. Moreover, during file processing, the annotators were in direct and constant contact with the leading team for questions and instructions. Finally, it is worth noting that most of the times the members of each

group were working together, as a team, and not individually.

5.1 Annotation Scheme

Figure 1 shows tier dependencies of the multi-tiered annotation scheme:

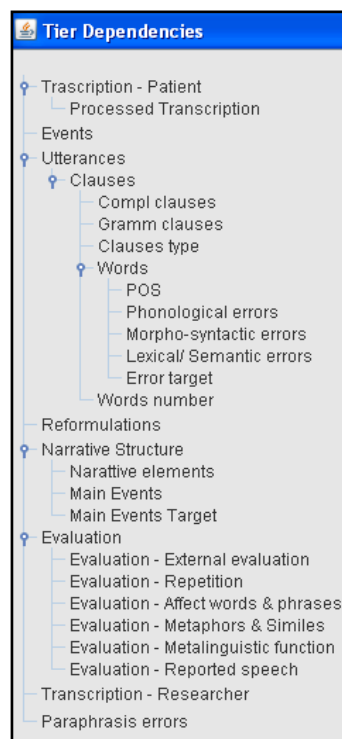


Figure 1: Tier dependencies of the annotation scheme

A set of detailed criteria regarding accurate transcription, definition of speech segment boundaries (utterances, clauses, and words), identification of each annotation category, and assignment of a valid value at each one were provided to the annotators. In the following sections, a brief description of the main annotation tiers is provided.

5.1.1. Patient transcription

This group includes two transcription tiers: the primary or “rough” transcription (parent tier) and the secondary, processed transcription (child tier). The first one contains anything which has been uttered by the participant, orthographically transcribed in the Greek alphabet. Processed transcription is the result of “cleaning up” the primary transcription of: a) repetitions: all but the final occurrence of a repeated word, phrase or segment were eliminated, excluding repetition for emphasis, b) self-corrections, c) formulaic phrases, d) one-word replies, e) parts of discourse irrelevant to the narrative content. Processed transcription provides the basis for the linguistic measurements of participants' discourse (number of utterances, clauses, words), as well as for the measurements of verbal flow (words / minute), verbal disruption, syntactic complexity, and narrative macrostructure.

5.1.2. Speech and non-speech events

In this tier the events of spoken discourse are annotated, such as vowel and consonant lengthening, silence, pauses (longer than 0.5 sec), noise, filled gaps, etc. This tier refers to the primary, “rough” transcription tier.

5.1.3. Reformulations

This tier contains: a) self-corrections at word level (e.g. “i podherta tis i p i mitria tis eklepse to dhaxtilidhi”, transl. “her (neologism targeting the word ‘stepmother’) her p her stepmother stole the ring”), which can be phonological, lexical or morphological; b) repetitions (e.g. “pire pire to to dhapidhoni”, transl. “he got he got the the (neologism targeting the word ‘ring’)”). Repetitions which are used for emphasis and serve an evaluative function in the narrative are not tagged as reformulations (e.g. “pias’ tin Eleni pias’ tin” transl. “catch Helen catch her”).

5.1.4. Utterances

This tier includes two child tier groups: clauses and words. The term “utterance” is used as equivalent to the term “sentence”, adopting the dominant view in linguistics that a sentence can consist of one or more clauses. The terms “utterance” and “sentence” are often used interchangeably in aphasia research (Faruqi-Shah & Thompson, 2007; Fyndanis, Varlokosta & Tsapkini, 2012; Wang, Yoshida & Thompson, 2014), following mainly Saffran, Berndt and Schwartz (1989: 471), who identify a set of certain structural types of sentences as utterances. Following Thompson et al. (1995), we used a combination of prosodic and semantic criteria to determine utterance boundaries. Utterance is defined as the speech section which follows and precedes silence, coincides with an intonational curve, and corresponds to a coherent meaningful unit of discourse. In cases where the aphasic speech was so fragmented that an intonational curve was difficult to identify, semantic criteria (coherence and completeness of meaning) were mainly used to define the utterance boundaries.

The tier of clauses is a child tier to the one of utterances. The presence of a verb was used to determine a clause. However, it should be noted that in aphasia verbs are prone to omission. Therefore, the presence of an overt subject or object (or of both) could also be employed as a sufficient criterion to identify a clause. Each clause was further annotated as:

- a) complete or incomplete: an incomplete clause is a clause that lacks some basic arguments or is abandoned before its meaning is completed. For example, the clause “ksafnika kapu pidhi egho... ox thee mu” (transl. “suddenly somewhere because I... oh my god”), was annotated as incomplete.
- b) grammatical or ungrammatical: ungrammatical is a clause which contains grammatical errors at word level or lacks basic arguments. For example, the clause “pefto to aftocinito” (transl. “I fall the car”) was annotated as ungrammatical, due to the omission of the preposition “apo” (transl. “from”).

Clauses were further annotated for their type. Values for clause types include independent clause, elliptical clause, as well as all types of subordinate clauses (clauses of time,

cause, result, purpose, conditional, relative, etc.) and verb complements.

The tier of words is a child tier to the one of clauses. Number of words in each clause is indicated and each word is further annotated with respect to its POS and to the phonological, morphosyntactic, and lexical/semantic errors it might contain.

Phonological errors are errors of phoneme omission (e.g. “cek” instead of “ceik”, transl. “cake”), substitution (e.g. “jelona” instead of “chelona”, transl. “tortoise”), addition (e.g. “setrono” instead of “strono”, transl. “spread”), etc.

Morphosyntactic errors are errors of omission (e.g. “laghos” instead of “o laghos”, transl. “hare” instead of “the hare”) or substitution (e.g. “theli pu pai” instead of “theli na pai”, transl. “wants that go” instead of “wants to go”) of free morphemes, as well as errors of agreement, such as number, case, and gender agreement between article and noun (e.g. “ton (def art masc) dhaxtilidhi (N neut)” instead of “to (def art neut) dhaxtilidhi”, transl. “the ring”), incorrect choice of aspect (e.g. “treksi” (pfv asp) instead of “trechi” (ipvf asp), transl. “runs”), tense (e.g. “pigha”, transl. “I went” instead of “pijeno”, transl. “I go”), case (e.g. “to ipe” (clit pro acc) instead of “tu ipe” (clit pro gen), transl. “told him”), etc.

Lexical/semantic errors include cases such as: a) neologisms, either retaining the morpho-phonological structure of Greek words (e.g. “dheklidhoni” instead of “dhaxtilidhi”, transl. “ring”), or not retaining it, so the word’s grammatical category is unspecified (e.g. “idhesofoliberi”, target word: unknown); b) production of words which have a phonological (e.g. “sidhora” instead of “simera”, transl. “today”) or semantic (e.g. “aschimi” instead of “omorfi”, transl. “ugly” instead of “pretty”) relationship with the target word.

Regarding word counting, the criteria proposed by Nicholas and Brookshire (1993) were followed: to be counted as words, lexical items have to be intelligible in context but not necessarily complete, accurate and relevant to the story. For example, the word “pipidhizis” which is a neologism (target word: “titivizis”, transl. “you chirp”) was counted as a word, even not phonologically accurate, while the word segment “che” (probably targeting the word “chelona”, transl. “tortoise”) was not counted as a word.

5.1.5. Narrative structure

This group of annotation tiers refers to the analysis of narrative discourse at the level of macrostructure. More specifically, it includes: a) a tier where the components of narrative structure are annotated (“narrative elements”) and b) a tier where the main informational units of discourse, the story’s “main events” are annotated (“main events”).

The structural components of the elicited narratives are annotated on the basis of the Labovian model of narrative structure (Labov, 1972; Labov & Waletzky, 1967), which includes the following structural units:

- a) Abstract: A single or multi-clause unit which informs the addressee on what the story is about (e.g. “theli na mas pi to paramithi ti ti chelona me to lagho”, transl. “the fairy tale wants to tell us (about) the tortoise with the hare”).

b) Orientation: The setting of the story, informing the addressee on the main characters (who), the place (where), and the time (when) of the story (e.g. “lipon i chelona ena proino itan sto dh dhasos”, transl. “well, in the morning the tortoise was in the woods”).

c) Complication: A sequence of events describing a ‘problem’, an unexpected complication for one or more of the main characters, his/their response(s) to the problem and his/their plan of action and attempts to resolve it. The events are leading to the climax or high point of the narrative.

d) Resolution: The part of the narrative which describes the outcome of the character(s)’ attempts to resolve the problem, leading to the narrative’s closure (e.g. “i chelona itane sto dhelos. ce itane medh me echi nicisi. ce exase o loghos o laghos”, transl. “the hare was at the end. and she was with, she has won. and the hare lost.”).

e) Coda: A unit linking the narrative to the present time (e.g. “And they lived happily ever after”).

Furthermore, for measuring the stories’ informational content, the number of “main events” was used as an indicator (Capilouto, Wright & Wagowich, 2006; Wright et al., 2005). Main events are defined as single or multi-clause units of a story, each one referring to a significant event of the story, which, at the same time, is independent of the other story events. Main events usually include one or more associated events and the temporal and causal relationships between them. Stories of tasks 2, 3 and 4 had a predefined number of main events. For example, the main events of the “hare and tortoise” story are the following:

1. The hare is going for a walk in the woods looking for food.
2. He meets the tortoise and thinks her slow walking is very funny.
3. The hare laughs at the tortoise and she challenges him to a race.
4. The hare finds her proposal very funny but he accepts the challenge.
5. They appoint the fox, the smartest animal, as the referee, and the race begins the following morning, when all the animals are gathered to watch it.
6. The hare decides to take a nap because he is confident that he can cover the distance from the tortoise very easily as soon as he gets up.
7. The tortoise keeps walking.
8. The hare sleeps for a little longer and when he wakes up he starts running.
9. He finds it strange not to see the tortoise anywhere but he thinks she gave up the race.
10. When the hare reaches the finishing line, he sees that the tortoise has finished first.

In order to be tagged as a main event, a textual unit should include the respective event or sequence of associated events as well as the relationship between them. For example, the following part of a story produced by a speaker with aphasia was tagged as main event no 1: “vjice o jo laghos ce vj vj vji vji vjice o laghos ce zjicise to xajito tou” (transl. “the hare went out and looked for his food.”).

5.1.6. Evaluation

The evaluation tier is not embedded in the narrative structure group of tiers, since evaluative devices might cross the boundaries of stories’ structural components. The category of evaluation includes linguistic devices which indicate the emotional and cognitive status of the narrator and his attitude towards the story events and characters. Evaluation expresses the narrator’s involvement in the narrative and constitutes a second narrative layer, which transforms a simple sequence of events to a worth-telling story. In line with previous studies on evaluation of aphasic discourse (Armstrong, 2005; Armstrong & Ulatowska, 2007, 2006; Ulatowska et al., 2006, 2011), the following evaluative features are annotated:

a) external evaluation: narrator’s comments, sometimes directly addressed to story recipient (e.g. “katalavenis?”, transl. “do you understand?”)

b) repetition of words or phrases for emphasis (e.g. “posa atoma itane, para pola itane”, transl. “so many people were there, a lot of people were there”)

c) words and phrases indicating emotional state, as well as inherently evaluative lexical items (e.g. “iche nevriasi”, transl. “he was upset”, “eftixos”, transl. “fortunately”)

d) metaphors and similes (e.g. “san salighari”, transl. “like a snail”)

e) metalinguistic function: narrator’s comments on his own speech (e.g. “edho itane... *dhe thimame*”, transl. “here there was... *I don’t remember*”, “laghos fajito chelona... *pos to len... to...*”, transl. “(the) hare (was looking for) food (the) tortoise... *how is it called... the...*”)

f) reported speech (direct and indirect) (e.g. “tha pas ghrighora ston aghona? tha pao ghrighora”, transl. “are you going to run fast at the race? I will run fast”).

6. Conclusion

The GREECAD is the first systematic attempt to develop an annotated corpus of aphasic discourse for Greek. The annotated transcripts of individuals with aphasia and healthy controls included in the current version of the corpus are being analyzed in terms of a set of measures, such as: a) verbal production and verbal flow (number of utterances, sentences and words, MLU, words/minute), b) syntactic complexity and grammaticality (number of conjunctions/total number of words, number of grammatical clauses/total number of clauses, number of subordinate clauses/total number of clauses, noun/verb ratio, number of errors/total number of words, etc.), c) verbal disruption: self-corrections, repetitions, abandoned clauses, gap-fillers, formulaic expressions, d) narrative structure (number of main events, narrative structure units, number of clauses/unit, evaluative devices by category, etc.). The main aims of the current studies being conducted or future ones are: a) to identify specific impairments at grammatical, lexical, and discourse level in the speech production of Greek-speaking individuals with aphasia, which could contribute to more effective evaluation, treatment, and assessment of treatment outcomes; b) to evaluate the overall communication abilities of speakers

with aphasia in their everyday lives, using narrative discourse as an objective communicative condition. Initial findings show differences in the group of individuals with aphasia compared to the control group regarding verbal production and flow, verbal disruption, grammatical accuracy, and syntactic complexity. However, narrative measures show a relative preservation of communication skills at the discourse level, since speakers with aphasia are able to produce the main informational content of a narrative and retain the main elements of narrative macrostructure despite their impairments at the microlinguistic level (Stamouli & Karasimos, 2015).

It is worth noting that the annotation scheme designed for the development of the GREECAD has been proven functional, flexible and broad, allowing the extended linguistic annotation of discourse samples in a consistent way.

An unrestricted online version of the GREECAD is not yet available. However, as soon as the ongoing studies of the research team are completed, free access to the fully annotated corpus with the XML metadata will be provided for research purposes and the corpus will be shared as a language specific resource to the META-SHARE¹ open-source repository.

The availability of the annotated aphasic discourse samples to the research community in combination with the rich metadata that accompany them, are expected to increase interest in the linguistic study of aphasia in Greek and to support the interdisciplinary study of aphasia, thereby contributing to a deeper and broader investigation of the complex phenomenon of aphasia.

7. Acknowledgements

This research action was co-financed by the European Union (European Social Fund) and Greek national funds through the operational program “Education and Lifelong Learning” of the National Strategic Reference Framework, research program THALES-UoA “Levels of impairment in Greek aphasia: Relationship with processing deficits, brain region, and therapeutic implications” (PI: Spyridoula Varlokosta).

We are grateful to Constantin Potagas, Ilias Papathanasiou, Georgia Kolintza, and Ioannis Evdokimidis for patient referral, to Margarita Atsidakou, Eva Efstratiadou, Anastasia Archonti, Ariadni Chatzidaki, Christoforos Routsis, Lina Chatziantoniou, and Dimitrios Kasselimis for clinical testing, to Athina Kontostavlaki for the development of the project Database, to Eleni Konsolaki for the coordination of the Annotators Team, to Aikaterini Pantoula, Pola Drakopoulou, and Sofia Apostolopoulou for the data collection, and to Christina Alexandri, Sofia Apostolopoulou, Dimitra Arfani, Kelly Atsali, Alexandros Bakogianis, Sofia Banou, Konstantina Daraviga, Vassiliki Dechounioti, Evi Doli, Pola Drakopoulou, Martha Drouga, Popi Eldahan-Apergi, Spyridoula Gasteratou, Mariana Georgouli, Aliko Gotsi, Nasia Gouma, Konstantina Goni, Chara Gourgouleti, Dimitris Katsimpokis, Giannis Kritikos,

Alexandra Kostaletou, Panayotis Kounoudis, Anna Michalaki, Sofia Pagoni, Panagiotis Panagopoulos, Kalliopi Papadodima, Aikaterini Pantoula, Stamatina Remoundou, Maria Saridaki, Alexandros Tsaboukas, and Ilias Valaskatzis for their valuable work in the transcription and annotation of the speech samples. Last, we would like to thank the six anonymous reviewers for their constructive comments, which helped us improve the paper.

8. Bibliographical References

- Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, 14, pp. 875-892.
- Armstrong, E. & Ulatowska, H. K. (2007). Making stories: Evaluative language and the aphasia experience. *Aphasiology*, 21, pp. 763-774.
- Armstrong, E. & Ulatowska, H. K. (2006). Stroke stories: Conveying emotive experiences in aphasia. In M. J. Ball & J. S. Damico (Eds.), *Clinical Aphasiology: Future Directions*. Hove, UK: Psychology Press.
- Armstrong, E. (2005). Expressing opinions and feelings in aphasia: Linguistic options. *Aphasiology*, 19, pp. 285-296.
- Berko-Gleason, J., Goodglass, H., Obler, L., Green, E., Hyde, M. & Weintraub, S. (1980). Narrative strategies of aphasics and normal-speaking subjects. *Journal of Speech and Hearing Research*, 23, pp. 370-382.
- Bird, S. & Liberman, M. (1999). A Formal Framework for Linguistic Annotation. Technical Report (MS-CIS-99-01). University of Pennsylvania.
- Capilouto, G. J., Wright, H. H. & Wagovich, S. A. (2006). Reliability of main event measurement in the discourse of individuals with aphasia. *Aphasiology*, 20, pp. 205-216.
- De Roo, E. (1999). *Agrammatic Grammar: Functional Categories in Agrammatic Speech*. Hague: Theuseus.
- Doyle, P. J., McNeil, M. R., Spencer, K. A., Goda, A. J., Cottrell, K. & Lustig, A. P. (1998). The effects of concurrent picture presentations on retelling of orally presented stories by adults with aphasia. *Aphasiology*, 12, pp. 561-574.
- Farqi-Shah, Y. & Thompson, C. K. (2007). Verb inflections in agrammatic aphasia: Encoding of tense features. *Journal of Memory and Language*, 56, pp. 129-151.
- Fyndanis, V., Varlokosta, S., & Tsapkini, K. (2012). Agrammatic production: Interpretable features and selective impairment in verb inflection. *Lingua*, 122, pp. 1134-1147.
- Harley, T. (2001). *The Psychology of Language: From Data to Theory* (2nd edition). New York: Psychology Press.
- Ide, N. & Suderman, K. (2007). GrAF: A graph-based format for linguistic annotations. In *Proceedings of the Linguistic Annotation Workshop*. Stroudsburg, PA: Association for Computational Linguistics, pp. 1-8.
- Ide, N. & Suderman, K. (2014). The linguistic annotation framework: A standard for annotation interchange and merging. *Language Resources and Evaluation*, 48, pp.

¹ URL: <http://www.meta-share.eu>

- 395-418.
- Kakavoulia, M., Stamouli, S., Foka-Kavalieraki, P., Economou, A., Protopapas, A. & Varlokosta, S. (2014). A battery for eliciting narrative discourse by Greek speakers with aphasia: Principles, methodological issues, and preliminary results [in Greek]. *Glossologia*, 22, pp. 41-60.
- Labov, W. (1972). *Language in the Inner City*. Philadelphia: The University of Pennsylvania Press.
- Labov, W. & Waletzky, J. (1967). Narrative analysis. In J. Helm (Ed.), *Essays in the Verbal and Visual Arts*. Seattle: University of Seattle Press, pp. 12-44.
- MacWhinney, B., Fromm, D., Forbes, M. & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, 25, pp. 1286-1307.
- MacWhinney, B., Fromm, D., Holland, A. & Forbes, M. (2012). AphasiaBank: Data and methods. In N. Mueller & M. Ball (Eds.), *Methods in Clinical Linguistics*. New York: Wiley, pp. 31-48.
- McNeil, M. R., Sung, J. E., Yang, D., Pratt, S. R., Fossett, T. R. D., Pavelko, S. & Doyle, P. J. (2007). Comparing connected language elicitation procedures in person with aphasia: Concurrent validation of the Story Retell Procedure. *Aphasiology*, 21, pp. 775-790.
- Menn, L., Ramsberger, G. & Helm-Estabrooks, N. (1994). A linguistic communication measure for aphasic narratives. *Aphasiology*, 8, pp. 315-342.
- Mesulam, M. M. (2000). *Principles of Behavioral and Cognitive Neurology* (2nd edition). New York: Oxford University Press.
- Nicholas, L. E. & Brookshire, R. H. (1995). Presence, completeness and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech and Hearing Research*, 38, pp. 145-156.
- Nicholas, L. E. & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, 36, pp. 338-350.
- Obler, L. K. & Gjerlow, K. (1999). *Language and the Brain*. Cambridge: Cambridge University Press.
- Olness, G. S. & Ulatowska, H. K. (2011). Personal narratives in aphasia: Coherence in the context of use. *Aphasiology*, 25, pp. 1393-1413.
- Papathanassiou, E., Papadimitriou, D., Gavrilou, V. & Michou, A. (2008). Normative data for the Boston Diagnostic Aphasia Battery in Greek: Gender and age effects [in Greek]. *Psychology*, 15, pp. 398-410.
- Saffran, E. M., Sloan-Berndt, R. & Schwartz, M. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language*, 37, pp. 440-479.
- Simos, P.G., Kasselimis, D. & Mouzaki, A. (2011). Age, gender, and education effects on vocabulary measures in Greek. *Aphasiology*, 25, pp. 492-504.
- Stamouli, S. & Karasimos, A. (2015). The Greek Corpus of Aphasic Discourse. Oral presentation at the workshop on the "Interdisciplinary Study of Aphasia" Thales Project. University of Athens. Athens, 27 June 2015.
- Thompson, C. K., Shapiro, L. P., Tait, M. E., Jacobs, B. J., Schneider, S. L. & Ballard, K. J. (1995). A system for the linguistic analysis of agrammatic language production. *Brain and Language*, 51, pp. 124-127.
- Ulatowska, H. K., Freedman-Stern, R., Doyel, A. W., Macaluso-Haynes, S. & North, A. (1983). Production of narrative discourse in aphasia. *Brain and Language*, 19, pp. 317-334.
- Ulatowska, H. K., North, A. J. & Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse in aphasia. *Brain and Language*, 13, pp. 345-371.
- Ulatowska, H. K., Olness, G. S., Keebler, M. & Tillery, J. (2006). Evaluation in stroke narratives: A study in aphasia. *Brain and Language*, Special Issue Academy of Aphasia 2006 Program, 99 (1-2), pp. 51-52.
- Ulatowska, H. K., Reyes, B. A., Santos, T. O. & Worle, C. (2011). Stroke narratives in aphasia: The role of reported speech. *Aphasiology*, 25, pp. 93-105.
- Varlokosta, S., Karasimos, A., Stamouli, S., Kakavoulia, M., Markopoulos, G., Goutsos, D., Fyndanis, V., Nerantzini, M. & Pantoula, A. (2013). Greek Corpus of Aphasic Discourse. Transcription and Annotation Guide [in Greek]. Athens: National and Kapodistrian University of Athens.
- Vermeulen, J., Bastiaanse, R. & van Wagensingen, B. 1989. Spontaneous speech in aphasia: A correlational study. *Brain and Language*, 36, pp. 252-274.
- Wang, H., Yoshida, M. & Thompson, C. K. (2014). Parallel functional category deficits in clauses and nominal phrases: The case of English agrammatism. *Journal of Neurolinguistics*, 27, pp. 75-102
- Westerhout, E. & Monachesi, P. (2006). A pilot study for a Corpus of Dutch Aphasic Speech (CoDAS). In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006), pp. 1648-1653.
- Williams, C., Thwaites, A., Buttery, P., Geertzen, J., Randall, B., Shafto, M., Devereux, B. & Tyler, L. (2010). The Cambridge Cookie-Theft Corpus: A corpus of directed and spontaneous speech of brain-damaged patients and healthy individuals. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010), pp. 2824-2830.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation, pp. 1556-1559. (URL: <http://tla.mpi.nl/tools/tla-tools/elan/>, Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands).
- Wright, H. H. (2011). Discourse in aphasia: An introduction to current research and future directions. *Aphasiology*, 25, pp. 1283-1285.
- Wright, H. H., Capilouto, G. J., Wagovich, S. A., Cranfill, T. & Davis, J. (2005). Development and reliability of a quantitative measure of adults' narratives. *Aphasiology*, 19, pp. 263-273.